Stability of scale quality under various data collection procedures: A mode comparison on the 'de Jong Gierveld Loneliness Scale'

van Tilburg, T.G.; de Leeuw, E.D.

published in
International Journal of Public Opinion Research
1991

DOI (link to publisher)
10.1093/ijpor/3.1.69

document version
Publisher's PDF, also known as Version of record

International Journal of Public Opinion Research Vol. 3 No. 1 0954–2892/91 $3.00

STABILITY OF SCALE QUALITY UNDER VARIOUS DATA COLLECTION PROCEDURES: A MODE COMPARISON ON THE 'DE JONG-GIERVELD LONELINESS SCALE*

Theo van Tilburg and Edith de Leeuw

ABSTRACT

Data collection procedures can influence respondents' self-disclosure, accuracy and motivation to complete the interview. In comparing research results across different studies, it is important to use robust measuring instruments. The 'De Jong-Gierveld Loneliness Scale' was developed to measure loneliness among different populations and in studies with different designs. Data on this loneliness scale were re-analyzed to investigate the robustness of the scale. The data were from six Dutch surveys. Different interview modes were used for data collection: three surveys with self-administered paper questionnaires, two surveys with face-to-face interviews, and one telephone survey. In order to compare the properties of the loneliness scale, a relatively homogeneous category of respondents was selected: single women between the ages of 25 and 65.

An examination of the scale with regard to five aspects of robustness showed in very few cases that it was affected. No evidence was found for the assumption that the use of a self-administered questionnaire would lead to high item non-response, any higher than

*This study was funded by grants from the Royal Dutch Academy of Sciences (KNAW) and the Social Cultural Sciences Foundation, which is subsidized by the Netherlands Organization for the Advancement of Scientific Research (NWO).

The authors wish to thank Dr J. J. Hox (University of Amsterdam) for statistical advice, Drs W. M. de Pijper (Vrije Universiteit Amsterdam) for writing a computer program, and Drs P. A. Dykstra (Vrije Universiteit Amsterdam) for help in the preparation of the manuscript. Furthermore, we are grateful to Prof Dr P. G. Swanborn, Dr M. P. M. de Goede, Dr W. Jansen and the members of the Department (especially Prof Dr J. de Jong-Gierveld, Prof Dr J. van der Zouwen, Dr W. Dijkstra, Drs P. A Dykstra and Drs J. Smit) and the anonymous reviewers for their helpful comments on an earlier draft. The authors gratefully acknowledge Dr M. P. M. de Goede, Dr G. H. Maassen (both University of Utrecht) and Prof Dr J. de Jong-Gierveld (Vrije Universiteit Amsterdam) (study 0), Drs H. Smitsmans (Valkenburg aan de Geul) (study 1), Dr L. R. Kooij (Zwammerdam) (study 4), and Drs L. J. van Doorn (NIPO, Amsterdam) (study 5) who kindly gave permission to re-analyze their data; Prof Dr J. de Jong-Gierveld (Vrije Universiteit Amsterdam), Prof Dr P. G. Swanborn, Dr J. Dessens, Dr W. Jansen (all University of Utrecht) and Dr J. J. Hox (University of Amsterdam) for their co-operation in the design of study 3, and Dr J. Dessens and Dr W. Jansen for organizing the data collection of study 3.

using other data collection procedures. It was also assumed that in self-administered questionnaires or telephone interviews, a better inter-item homogeneity and a better person scalability would be found in studies with face-to-face interviews. The results sustained this hypothesis. Further, it was believed that the absence of an interviewer would result in greater self-disclosure and therefore in higher scale means. We found no evidence to support this.

In general, the results showed that the loneliness scale met the psychometric requirements of items non-response, scale homogenity and person scalability. After testing the robustness of the scale, we conclude that it is questionable on two aspects: the inter-item homogeneity and the person scalability.

INTRODUCTION

Until recently, the face-to-face interview was the favorite data collection method in social surveys. However, the high costs and growing non-response of face-to-face interviews have led survey researchers to consider 'alternative' data collection procedures such as mail surveys with self-administered paper questionnaires and telephone interviews (Dillman, 1978; Groves and Kahn, 1979; Van der Zouwen and De Jong-Gierveld, 1987). Further recent developments in survey methodology are computer-assisted face-to-face interviews (with an interviewer), computer-assisted self-administered interviews (without an interviewer) (CBS-select, 1987; Nicholls and Groves, 1986) and mixed-mode designs (Dillman and Tarnai, 1988).

In using new data collection procedures (modes) and especially in combining different modes in one design, it is of paramount importance to know how data collection procedures can influence research results. For that reason, research on mode comparisons is increasing both in the United States and in Europe. In a review of the published literature on this topic, face-to-face interviews resulted in slightly better overall data quality as compared with telephone interviews (De Leeuw and Van der Zouwen, 1988), whereas self-administered questionnaires in comparison to both face-to-face and telephone interviews resulted in slightly better data, especially if more privacy-sensitive questions were asked (De Leeuw *et al.*, 1989). This poses a problem for the comparability of survey results.

One way to solve this problem is to develop robust research instruments. De Jong-Gierveld has developed a potentially robust 11-item scale for measuring the concept of loneliness in face-to-face interviews, self-administered interviews and other modes (De Jong-Gierveld and Van Tilburg, 1986).

The Development of the Loneliness Scale

The conceptualization of loneliness drew upon the cognitive approach to loneliness (Perlman and Peplau, 1981). Within that approach, loneliness was a

subjective experience and, as such, was not directly related to situational factors. The importance of social perceptions and evaluation of one's personal relationships was emphasized. Loneliness or 'subjective social isolation' was a condition in which a person experienced a distressing or unacceptably low quantity and/or quality of personal relationships. This description included situations where the number of existing relationships was smaller than was considered desirable or acceptable, as well as situations where the desired intimacy had not been realized (De Jong-Gierveld, 1989).

Originally, a 34-item multi-dimensional scale of loneliness (De Jong-Gierveld and Raadschelders, 1982) was developed. This scale consisted of several subscales. Of these subscales, the most central to loneliness was a 9-item deprivation scale. In developing the scale, the researchers started with a content analysis of compositions written by 114 people about their loneliness experiences; these people were approached via singles clubs and pastors. Next, in step 2 of the development process, items derived from the compositions were tested in a pilot study of 59 women and men. A revised set of items was included in a questionnaire which was administered by means of semi-structured face-to-face interviews with 556 women and men (step 3). Since the 9-item deprivation scale was primarily found to measure severe feelings of loneliness, changes were made. In step 4, 30 items were selected, beginning with the 9 items of the deprivation scale. Additional items were obtained from answers to open-ended questions (which were part of the interview in step 3) in which people expressed their loneliness. Out of these 30 items, an 11-item unidimensional scale (Figure 1) was developed, using data of unemployed, disabled, and employed men and women. The data were gathered by a self-administered questionnaire presented at the end of a face-to-face interview. The scale (1) assessed severe feelings of loneliness as well as less intense loneliness feelings, (2) consisted of negative as well as positive items, and (3) represented a latent continuum of deprivation. In addition, the scale met the criteria of the dichotomous logistic Rasch model[1] (De Jong-Gierveld and Kamphuis, 1985).

The Purpose of this Study

The performance of the 11-item scale in four other research projects (using self-administered paper questionnaires as well as face-to-face interviews) was investigated (De Jong-Gierveld and Van Tilburg, 1987). The results showed

[1] The Rasch model is designed for dichotomous variables, whereas the latent trait is assumed to be continuous. The four assumptions underlying the Rasch model are (1) unidimensionality, (2) local stochastic independence, (3) monotonicity, and (4) sufficiency of simple sum statistics. The majority of the goodness-of-fit tests of the Rasch model are based on invariance of the item difficulties over samples. The invariance of item difficulties must also hold if the sample is divided on the basis of characteristics such as race, sex, or test score profiles (for an introduction of the Rasch model, see Van den Wollenborg, 1979).

FIGURE I **The 'De Jong-Gierveld Loneliness Scale'** (De Jong-Gierveld and Kamphuis, 1985)

*Scale items**

1.	There is always someone that I can talk to about my day to day problems	yes! yes	more-or-less	no no!
2.	I miss having a really close friend	yes! yes	more-or-less	no no!
3.	I experience a general sense of emptiness	yes! yes	more-or-less	no no!
4.	There are plenty of people that I can lean on in case of trouble	yes! yes	more-or-less	no no!
5.	I miss the pleasure of the company of others	yes! yes	more-or-less	no no!
6.	I feel my circle of friends and acquaintances is too limited	yes! yes	more-or-less	no no!
7.	There are many people that I can count on completely	yes! yes	more-or-less	no no!
8.	There are enough people that I feel close to	yes! yes	more-or-less	no no!
9.	I miss having people around	yes! yes	more-or-less	no no!
10.	Often, I feel rejected	yes! yes	more-or-less	no no!
11.	I can call on my friends whenever I need them	yes! yes	more-or-less	no no!

*The sequence of the items in the studies 0, 1, 2 and 4 was: 2, 5, 11, 10, 3, 7, 9, 1, 4, 6, 8. No show cards were used in the face-to-face interviews.

Processing the scale data

The five-category responses were transformed into dichotomous responses. Responses indicating a (certain) feeling of loneliness were assigned a score of one loneliness point. That is, if the response 'more or less', 'yes', or 'yes!' was given to a negatively formulated item (item numbers 2, 3, 5, 6, 9, 10) *or* if the response 'no!', 'no', or 'more or less' was given to a positively formulated item (item numbers 1, 4, 7, 8, 11), a scale point was assigned. According to this procedure, the 'more or less' answers were not viewed as neutral answers, but as indicators of loneliness. The other answers were assigned a zero score. Thus, in the case of extreme loneliness, a respondent could score a total of 11 loneliness points. The minimum score was 0.

If a respondent had scored *one and only one* missing value, the response was not viewed as a loneliness indicator; thus a zero score was given for the item. If a respondent had scored two or more missing values, the particular case was deleted from the analysis.

that the construct validity and reliability of the scale in each of the research projects was sufficient. Furthermore, there was the impression (an impression which was not based on adequate analyses) that different models of data collection did not influence the mean score of the scale. However, adequate comparisons between the studies were not made, and therefore the robustness of the loneliness scale was tested in the present study.

The concept of data quality is a complex one (Bailar, 1984; Groves, 1989). As a consequence, multiple indicators of data quality have been used in studies on mode comparisons (cf. De Leeuw and Van der Zouwen, 1988; De Leeuw *et al.*, 1989). Authors had emphasized the extreme importance of the following indicators: low item non-response (Groves and Kahn, 1979; Groves, 1989; Dillman, 1978; Hochstim, 1967), scale reliability and inter-item variance (Aneshensel *et al.*, 1982; Herman 1977; Rogers 1976), and similarity of answer distributions and descriptive statistics (Colombotos, 1965; Hochstim, 1967; Körmendi and Noordhoek, 1989; Sykes and Collins, 1988; Sykes and Hoinville, 1985; Siemiatycki, 1979). Therefore we defined robustness as invariance of (a) item non-response, (b) inter-item (scale) homogeneity, (c) person scalability, (d) item p-values and (e) scale means under different data collection procedures.

Previous studies on mode effects noted small but worrisome differences in 'data quality' when different data collection procedures were compared. It is generally assumed that the presence or absence of an interviewer is an important factor in the process (Cannel and Fowler, 1963; Hyman, 1954; Sudman *et al.*, 1965).

For instance, the presence of an interviewer may influence the responses in a positive way. A skilled interviewer can use probes, explain a question, or use non-verbal communication to motivate respondents (Galtung, 1967; Sykes and Hoinville, 1985). This will induce the respondents to answer all the questions. Using self-administered paper questionnaires, it is easier for respondents to skip questions they find (emotionally) difficult, leading to a higher percentage of item non-response. However, the computer-assisted self-administered interview has an interactive component, that might increase respondent involvement beyond that evoked by paper questionnaires (Kiesler and Sproull, 1986). Furthermore, the internal control of the survey or interview program can prevent respondents from inadvertently skipping questions, it can perform checks on the answers or offer probes, explanations and general help (Sikkel, 1987), all of which lead to a minimum item non-response.

Galtung (1967) noted that in self-administered questionnaires, it is the respondent, and not the interviewer, who writes down the answers. This will provide the respondent with an extra check on the correctness and consistency of an answer, which may result in a higher inter-item homogeneity of the scale and in fewer deviant individual response patterns (De Leeuw and Hox, 1988; Tatsuoka and Tatsuoka, 1983; Van der Flier, 1982).

One of the main advantages of self-administered questionnaires is that the absence of the interviewer may introduce a greater feeling of anonymity in the respondent (Cannel and Fowler, 1963). The more anonymous and private setting where self-administered questionnaires are completed, reduces the tendency of respondents to present themselves in a favorable light (Ellis, 1947; Sudman and Bradburn, 1974). In general, compared to face-to-face interviews, self-administered questionnaires may present fewer problems of self-presentation, resulting in greater self-disclosure and more acknowledgements of loneliness feelings (cf. Hochstim, 1967; Wiseman, 1972; Siemiatycki, 1979).

The above considerations led us to investigate item non-response, inter-item homogeneity, person scalability, and item and scale means of the loneliness scale in studies using different modes. Data were available from Dutch loneliness studies using the following modes: face-to-face interviews, self-administered paper questionnaires and self-administered home computer-assisted interviews. The objectives of these studies, the populations studied, and the designs were heterogeneous. In order to make appropriate comparisons, especially with regard to the scale mean and the item p-values, the analyses were performed on a relatively homogeneous category of respondents: single women between the ages of 25 and 65.[2]

METHOD

The Data Sets

The loneliness scale was developed in a study conducted in 1982. Face-to-face interviews among a sample of 1702 unemployed, occupationally disabled and employed individuals and their partners were conducted (De Goede and Maassen, 1988). The sample was randomly selected from the national population, stratified according to employment status. The response rate of the face-to-face interviews was 74.4 per cent for the unemployed without partner ($N=258$),[3] 69.6 per cent for the unemployed with partner ($N=332$), 93.3 per cent for their partners ($N=231$),[4] 73.3 per cent for the occupationally disabled without parter ($N=210$), 76.4 per cent for the occupationally disabled with partner ($N=311$), 92.3 per cent for their partners ($N=287$), and 67.7 per cent for the employed individuals ($N=458$). At the end of the interview, a paper questionnaire was presented with a request to mail it to the researchers (De

[2] The choice of this category was determined by opportunity: it was the common category available in most of the loneliness studies.

[3] The unemployed and disabled respondents were selected with the help of an omnibus enquete. The members of the household were asked if the unemployed or disabled person in the household would like to co-operate with an interview.

[4] The partners of the unemployed and the occupationally disabled were only asked to co-operate if the unemployed and the occupationally disabled co-operated.

Jong-Gierveld and Kamphuis, 1985). The response rate of the additional mail questionnaires was 74.2 per cent ($N = 1702$). If the respondents had a partner, both individuals were requested to fill in the questionnaire. The central theme of the additional questionnaires was the interrelation between the degree of confidence and intimacy of achieved relationships and loneliness, thereby comparing individuals who were unemployed or occupationally disabled to individuals who were employed. The subsample for the present study consisted of 117 single women from 25 to 64. *Study 0* refers to these respondents. Data from these women will be used as a 'base-line' for factual comparison.

Study 1 was conducted in Maastricht in 1985 (Smitsmans, 1986). Self-administered paper questionnaires were mailed, using Dillman's total design method (Dillman, 1978), to a sample of non-married people. The sample was selected at random from the population register. 400 useful questionnaires were received (response rate 40.4 per cent). Out of these, 196 were from single women from 25 to 64. This particular study examined whether going to singles' clubs helped resolve loneliness.

Study 2, which was completed in 1986, involved face-to-face interviews with 419 adult men and women (Van Tilburg, 1988). The 31 interviewers were selected on the basis of their motivation for the research topic, their age (at least 25), and their training in face-to-face interviewing. They received detailed instructions about expected interviewer behavior. The sample was stratified according to marital status; equal numbers of married, never married, divorced and widowed men and women were selected at random from the population registers of Purmerend, a rapidly growing commuter city (population about 50,000) and Haarlemmermeer. The latter community includes 16 villages, the largest of which has a population of about 4,500, while the smaller ones have populations of approximately 400. The response rate was 47.7 per cent. As a result of the stratification criteria, a relatively large population of the respondents, 84 out of 419, were single women from 25 to 64. The objective of this study was to analyse the effects on loneliness of the support provided by the network of relationships.

Study 3 was conducted in and around the town of Utrecht in 1987. A number of streets were assigned at random to 60 interviewers, and from a certain point addresses were randomly selected. Undergraduate students performed the interviews as part of a course requirement. In face-to-face interviews, 294 adult men and women were questioned about loneliness and personal relationships (Jansen *et al.*, 1990). No information is available about non-response. Among the respondents 40 were single women from 25 and 65. In studies 2 and 3, no show cards were used when administering the questions on loneliness.

Study 4 was conducted among patients of general practitioners in the western part of Holland. The study was concerned with the number of visits and the reasons for consulting the general practitioner (Kooij, 1988). For the sake of convenience, 18 general practitioners were selected and asked to give all the addresses of single parents; 11 gave permission (61.8 per cent response). 173 single parents returned the mail questionnaire (79 per cent response); 148 of them were single women from 25 and 65.

In *study 5*, an electronic survey method was used. In this case, a panel of respondents answered questions by means of a computer (owned by a marketing research institution, and installed in respondents' homes). The questionnaires and the data were transported via a regular telephone connection. This procedure is called teleinterviewing (De Leeuw, 1988; Saris *et al.*, 1987). The panel was representative of the Dutch population (Van Doorn, 1987). In 1988, the women of the panel were asked to answer questions about individualization, personal relationships and loneliness. Among these respondents, 186 were single women.[5]

Procedure

For each survey, the mean proportion of item non-response on the loneliness scale was computed. Furthermore, for each survey the *p*-values (percentage of respondents who agreed with negatively formulated items or disagreed with positively formulated items) of the items, the scale mean (range 0–11; see Figure 1), KR-20, the item-rest correlations, and Loevinger's Homogeneity-coefficient (as proposed by Mokken) were computed. For each respondent with a loneliness scale score of 1 to 10, Van der Flier's Q was computed. KR-20 is an indicator of the reliability and internal consistency of dichotomous data. It is comparable with Cronbach's coefficient alpha.

Mokken (1971) developed a non-parametric item response theory, based on item characteristics. A Mokken scale is a probabilistic generalization of Guttman's scalogram (Sijtsma, 1988). Loevinger's H is often used to evaluate a scale in terms of Mokken scalability; H_i to evaluate an item in a scale. When the very strict Guttman scalogram model holds, H has the maximum value of 1. On the other hand an H-value of zero indicates an average covariance of zero between the items. According to Mokken, a scale can be ordered from weak to strong using the H-coefficient. A weak scale is indicated by $.30 \leqslant H < .40$; a strong scale by $H \geqslant .50$. H_is must be $\geqslant .30$.

Van der Flier's Q (Van der Flier, 1980) is useful to determine deviant

[5] The distribution by age was not given by the principal researcher and therefore not available in this secondary analysis.

individual response patterns (person inscalability). If, for example students answer 8 out of a total of 10 items correctly, one expects that they will have missed the two most difficult items. If instead, the two easiest items are answered incorrectly, their item response pattern is completely unexpected. Between the totally deviant and the perfect pattern, there is a wide range of different item response patterns. Q is the probability of a score pattern, given the p-values of the items within the conditional distribution of pattern probabilities (chosen by the researcher or derived from data). Q indicates the probability of a certain individual response pattern or an even more deviant response pattern, given the test score of the respondent, and can be interpreted as an ordinary one-tailed probability level (p must be $\geqslant .05$). In order to measure the degree of deviation, several indices of person scalability have been developed (Sijtsma, 1988; Van der Flier, 1980). Such indices can be useful diagnostic tools if they are used in addition to summary statistics, such as the total test score. Levine and Rubin (1979), for example, discussed the use of 'appropriateness indices' for the detection of cheating on aptitude tests, while Tatsuoka and Tatsuoka (1982; 1983) demonstrated the usefulness of their 'individual consistency index' in detecting students who used a wrong algorithm in problem-solving tasks. Person scalability indices can also be used to investigate the effect of various testing conditions (Van der Flier, 1980) or, as in our case, the influence of modes of data collection. For each respondent (with a loneliness scale score of 1 to 10) within each survey, the Q-value was computed, given the respondent's test scores and given the p-values of the items in study 0.

Analyses of variance were performed to compare the number of missing values, the mean Q-value, item p-values and the scale mean between six studies. The Mokken Test program is used to compare the H-values of the studies. It is conceivable that differences between the surveys can be attributed to other factors than to inter-method differences. For instance, the samples may differ in age and marital status. These potential differences can depress the robustness of the scale rather than the inter-method differences. Therefore, subsequent ANOVA-analyses were performed to investigate the effect of sample composition (age, marital status) on scale mean and Q-scores.

RESULTS

The psychometric properties of the loneliness scale in the six studies are shown in Table 1. The mean number of missing values on the 11 loneliness items (range 0–11) was between .01 and .28.[6] The relatively high means in studies 0 and 3

[6] The principal researcher of study 1 recoded the missing values to substitutes, e.g. the mean score for the entire sample of respondents. The procedure used in study 5 excluded missing values.

TABLE 1 **Psychometric Properties of the Loneliness Scale in Six Studies, with Self-administered Paper Questionnaires (sap), Face-to-Face Interviews (ftf), and Teleinterviews (ti)**

Study	0	1	2	3	4	5	
Mode	sap	sap	ftf	ftf	sap	ti	
N	117	196	84	40	148	186	
M#miss. val.	.15		.01	.28	.03		$F = 1.7$
SD#miss. val.	.86		.11	1.74	.16		
KR-20	.90	.89	.83	.80	.90	.86	
$r_{1,rest}$	.58	.50	.35	.34	.50	.51	
$r_{2,rest}$	.50	.65	.44	.51	.50	.54	
$r_{3,rest}$	.68	.60	.54	.59	.60	.61	
$r_{4,rest}$	.61	.73	.58	.47	.65	.52	
$r_{5,rest}$	.67	.58	.58	.46	.69	.56	
$r_{6,rest}$	.63	.65	.44	.42	.66	.60	
$r_{7,rest}$	.61	.59	.39	.31	.65	.50	
$r_{8,rest}$	.66	.60	.65	.44	.60	.58	
$r_{9,rest}$	.65	.68	.39	.51	.72	.61	
$r_{10,rest}$	.74	.57	.64	.66	.67	.51	
$r_{11,rest}$	.61	.58	.63	.39	.67	.58	
Mokken's H	.52	.49	.38	.34	.52	.43	$t = 12.3$ $p < .05$
H_1	.46	.40	.26	.26	.40	.38	$t = 5.5$
H_2	.39	.49	.32	.35	.46	.40	$t = 6.5$
H_3	.56	.51	.38	.46	.54	.47	$t = 5.4$
H_4	.54	.55	.46	.45	.52	.41	$t = 4.8$
H_5	.52	.45	.41	.32	.52	.41	$t = 6.4$
H_6	.52	.50	.31	.30	.51	.44	$t = 9.7$
H_7	.69	.53	.30	.24	.56	.51	$t = 21.7$ $p < .001$
H_8	.51	.46	.46	.31	.57	.42	$t = 8.7$
H_9	.54	.53	.29	.35	.59	.44	$t = 14.4$ $p < .05$
H_{10}	.58	.51	.48	.44	.52	.41	$t = 6.2$
H_{11}	.50	.45	.55	.28	.52	.44	$t = 6.3$
Study	0	1	2	3	4	5	
Mode	sap	sap	ftf	ftf	sap	ti	
N	117	196	84	40	148	186	
M Flier's Q	.50[a]	.34[bc]	.24[c]	.26[cd]	.37[bd]	.35[bd]	$F = 4.5$ $p < .001$
SD Flier's Q	.36	.35	.30	.33	.35	.34	
p_1	.33	.33	.32	.21[a]	.39[b]	.39[b]	$F = 1.3$
p_2	.32	.37	.40	.28	.43	.40	$F = 1.1$
p_3	.29[a]	.45[b]	.43[a]	.38	.51[b]	.42[b]	$F = 3.0$ $p < .05$

(continued)

p_4	$.39^{ad}$	$.36^{qd}$	$.26^{abc}$	$.13^{b}$	$.42^{d}$	$.30^{a}$	$F=3.5$ $p<.01$
p_5	$.31^{a}$	.38	.44	$.28^{a}$	$.46^{a}$	.36	$F=1.9$
p_6	.35	.38	.40	.31	.42	.38	$F=0.5$
p_7	$.55^{ac}$	$.48^{c}$	.46	$.33^{bc}$	$.55^{ac}$	$.58^{a}$	$F=2.3$ $p<.05$
p_8	$.34^{a}$	$.36^{a}$	.34	$.23^{a}$	$.47^{b}$	$.36^{a}$	$F=2.1$
p_9	$.26^{a}$	.34	$.44^{b}$	.31	.33	$.38^{b}$	$F=1.7$
p_{10}	$.29^{a}$	$.27^{a}$	$.30^{a}$	$.26^{a}$	$.43^{b}$	$.30^{a}$	$F=2.4$ $p<.05$
p_{11}	.30	.33	$.21^{a}$	.23	$.39^{b}$	.32	$F=1.8$
M Loneliness	3.7^{ab}	4.1	4.0	2.9^{a}	4.8^{c}	4.2^{bc}	$F=2.3$ $p>.05$
SD Loneliness	3.6	3.6	3.2	2.8	3.8	3.4	

[abcd]The difference between means with different superscripts is significant at $p<.05$ (LSD range test).

were each caused by a single respondent with 9 and 11 missing values, respectively. The differences were not significant.

KR-20, indicating the internal consistency of the loneliness scale, was reasonably good in all the surveys ($\geq .80$). In studies 2 and 3 (both face-to-face interviews) and study 5 (teleinterviews), KR-20 of the scale was lower in the three other studies (all self-administered paper questionnaires). All the item-rest correlations were sufficiently high ($\geq .30$).

Mokken's H in the various surveys indicated a weak scale in studies 2 and 3 (both face-to-face interviews), a moderate scale in study 1 (self-administered paper questionnaires) and in study 5 (teleinterviews), and a strong scale in studies 0 and 4 (both self-administered paper questionnaires). When the Mokken test was performed, the differences between the Homogeneity coefficients were found to be significant. As Table 1 shows, the Homogeneity coefficients of the individual items related to the scale (H_i) were not all above the critical level of .30 in studies 2 and 3 (both face-to-face interviews). The H_7 and the H_9 differed significantly in the studies, with again the lowest values for studies 2 and 3.

The mean of Van der Flier's Q for study 0 was .50. It was .50 because this study served as the basis for across-study comparisons. The mean of Van der Flier's Q in all the other studies was lower, indicating more deviant individual response patterns, given the test score of the respondents and the *p*-values of the items of study 0. The mean Q of studies 1 and 4 (both self-administered paper questionnaires) and study 5 (teleinterviews) was significantly lower than in study 0. The lowest mean Q was found for studies 2 and 3 (both face-to-face interviews). However, in these two studies the mean Q was also above the critical level of .05.

The results with regard to the *p*-values of the items and the scale means were the following. Significant differences between the *p*-values obtained in the six

TABLE 2 **Mean Q-value and Loneliness Score in Four Studies, with Age and Marital Status as Covariates**

	Q			*Loneliness Score*		
	N	*Unadjusted*	*Adjusted*	*N*	*Unadjusted*	*Adjusted*
1 (self-administered)	142	.34	.33	196	4.1	4.1
2 (face-to-face)	66	.24	.25	84	4.0	4.0
3 (face-to-face)	28	.25	.22	39	3.0	3.1
4 (self-administered)	105	.36	.38	146	3.6	3.7
		$F_{(3)} = 2.9$			$F_{(3)} = 2.0$	
		$p = < .05$			$p = > .05$	

studies were found for 9 of the 11 items. Significant differences between the mean scale scores were also found. In particular, the mean scale score of study 4 (self-administered paper questionnaires) was significantly higher then in study 3 (face-to-face interviews) and in study 0 (the principal survey, self-administered paper questionnaires). The mean scale score in study 5 (teleinterviews) was significantly higher than that in study 3.

As regards four studies, two with face-to-face interviews (2 and 3) and two with self-administered paper questionnaires (1 and 4), the question was posed as to whether the observed differences in the average Q-value and loneliness score could be attributed to differences in the composition of the samples. There also appeared to be differences between the various studies as to the age and marital status of the respondents. Young people, for example, were greatly overrepresented in study 3, as were older people in study 1. The unmarried were overrepresented in studies 1 and 4, as were the divorced in studies 2 and 4. There were hardly any widows in study 4 (Table 1).

ANOVA-analyses were performed to see whether the studies differed in mean Q-value and scale mean if age and marital status were held constant. If age and marital status were used as covariates, on the average a statistically significantly low Q-value, indicating more deviant individual response patterns, was noted in studies 1 and 3 (Table 2). The covariate marital status was significant: widows had relatively low Q-values. If age and marital status were used as covariates, the differences in mean loneliness scores were no longer significant (Table 2).

SUMMARY AND DISCUSSION

When comparing research results across different studies, it is important to use robust measuring instruments. The 'De Jong-Gierveld Loneliness Scale' was

developed with the purpose of measuring loneliness among different populations and in studies with different designs. In this article, data on this loneliness scale were re-analyzed to investigate the robustness of the scale. The data were from six surveys among different populations, with different topics, and with different designs. Different modes were used for data collection: three surveys with self-administered paper questionnaires (including the principal survey in which the scale was developed), two surveys with face-to-face interviews, and one survey with teleinterviews. In order to compare the properties of the loneliness scale, a relatively homogeneous category of respondents was selected: single women from 25 and 65. However, the results of the various studies showed considerable variability in loneliness within this category.

A comparison with regard to five aspects of robustness revealed a number of differences between the six studies. The question is whether the observed differences were related to the use of different data collection procedures, as suggested by De Leeuw *et al.* (1989).

No evidence was found for the assumption that the use of a self-administered questionnaire would lead to high item non-response, higher than with other data collection procedures. The differences in item non-response between the studies were not significant.

It was also suggested that in self-administered questionnaires or teleinterviews, a better inter-item homogeneity and a better person scalability would be noted than in studies with face-to-face interviews. In keeping with this idea, it was noted that the inter-item homogeneity and the person scalability of the loneliness scale in studies 2 and 3 (face-to-face interviews) were the weakest. It appears to be acceptable (in research among this selection of respondents) to use the loneliness scale in self-administered paper questionnaires or teleinterviews.

A potential tool to improve the performance of the face-to-face interview on this point could be visual aids. During the face-to-face interviews (studies 2 and 3), interviewers did not present show cards for the loneliness items, but in the self-administered procedures (including the teleinterview), the respondents were offered all the permitted response categories after each question. Sykes and Collins (1989) suggest that visual cues aid respondents to adequately differentiate between response categories that might sound familiar and encourage them to give a more considered reaction to repetitive questions. It is inconceivable that the absence of visual aids in the response process causes more (random) errors and lower inter-item homogeneity and person-scalability in studies 2 and 3. Further empirical research is needed to test this hypothesis. Meanwhile, it seems wise to use show cards in utilizing the loneliness-scale, especially in mixed-mode research.

It was assumed that the absence of an interviewer would result in greater self-disclosure and therefore in higher scale means. We did not find supporting

evidence for this hypothesis. In one study with self-administered paper questionnaires, a relatively high mean score was noted, and in one study with face-to-face interviews a relatively low mean score. However, the differences between the studies were not significant if differences in sample composition as to age and marital status were held constant.

In general, the results showed that the loneliness scale met the psychometric requirements of item non-response, scale homogeneity and person scalability. After testing the robustness of the scale, we can conclude that the robustness is questionable on two aspects: the inter-item homogeneity and the person scalability. However, this conclusion is based on data from only one relatively homogeneous category of respondents. Furthermore, it is based on a limited number of studies. Among the studies, there was considerable variability in topic, design and sample composition. The two face-to-face surveys also used different interviewers. To investigate the influence of data collection procedures on the quality of data, additional research is needed. A study with an experimental design is now in progress by the second author.

REFERENCES

Anashensel, C. S. *et al.* (1982): 'Measuring Depression in the Community: A Comparison of Telephone and Personal Interviews', *Public Opinion Quarterly*, **46**, 110–21.

Bailar, B. (1984): 'The Quality of Survey Data: *Proceedings of the Section on Survey Research Methods of the American Statistical Association.*

Cannel, C. F. and Fowler, F. J. (1963): 'Comparison of a Self-Enumerative Procedure and a Personal Interview: A Validity Study', *Public Opinion Quarterly*, **27**, 250–64.

CBS-select 4 (1987): 'Automation in Survey Processing'. The Hague: Staatsuitgeverij.

Colombotos, J. (1965): The Effect of Personal versus Telephone Interviews on Socially Acceptable Responses', *Public Opinion Quarterly*, **29**, 457–8.

de Goede, M. P. M. and Maassen, G. H. (1988): 'Beleving van Niet-Werken [The Experience of Being Jobless]', Doctoral Dissertation, University of Utrecht.

de Jong-Gierveld, J. (1989): 'Personal Relationships, Social Support and Loneliness', *Journal of Social and Personal Relationships*, **6**, 197–221.

de Jong-Gierveld, J. and Kamphuis, F. (1985): 'The Development of a Rasch-Type Loneliness Scale,' *Applied Psychological Measurement*, **9**, 289–99.

de Jong-Gierveld, J. and Raadschelders, J. (1982): 'Types of Loneliness'. In: L. A. Peplau and D. Perlman (eds.), *Loneliness: A Sourcebook of Current Theory, Research and Therapy*, New York, Wiley, pp. 105–19.

de Jong-Gierveld, J. and van Tilburg, T. G. (1986): *Manual of the Loneliness Scale.* (Available from the authors).

de-Jong-Gierveld, J. and van Tilburg, T. G. (1987): 'Het Meten van Persoonlijke Ervaringen en Gevoelens in Vragenlijstonderzoek: Een Studie naar het Functioneren van de Eenzaamheidsschaal in Verschillende Onderzoekingen [Measuring Personal

Experiences and Emotions with Questionnaires: A Study of the Properties of the Loneliness Scale in Different Research Projects]', In: J. de Jong-Gierveld and J. van der Zouwen (eds.), *De Vragenlijst in Sociaal Onderzoek: Een Confrontatie van Onderzoekspraktijk en-methodiek*, Deventer, Van Loghum Slaterus, pp. 67–83.

de Leeuw, E. D. (1988): 'Gegevens Verzamelen met behulp van een Computer [Computer Assisted Data Collection]'. In: J. J. Hox and G. de Zeeuw (eds.), *De Microcomputer in Sociaal Wetenschappelijk Onderzoek*, Lisse, Swets and Zeitlinger, pp. 55–72.

de Leeuw, E. D. and Hox, J. J. (1988): 'Artifacts in Mail Surveys: The Influence of Dillman's Total Design Method on the Quality of the Responses'. In: W. E. Saris and I. N. Gallhofer (eds.), *Sociometric Research*, London, MacMillan, Vol. 2, pp. 61–73.

de Leeuw, E. D. and van der Zouwen, J. (1988): 'Data Quality in Telephone and Face-to-face Surveys: A Comparative Meta Analysis'. In: R. M. Groves *et al.* (eds.), *Telephone Survey Methodology*, New York, Wiley, pp. 283–300.

de Leeuw, E. D. *et al.* (1989): 'Data Quality in Mail Surveys as Compared to Face-to-face and Telephone Interviews: A Meta-Analysis of the Research Literature', *Kwantitatieve Methoden*, **31**, pp. 123–46.

Dillman, D. A. (1978): *Mail and Telephone Surveys: The Total Design Method*, New York, Wiley.

Dillman, D. A. and Tarnai, J. (1988): 'Administrative Issues in Mixed Mode Surveys'. In: R. M. Groves *et al.* (eds.), *Telephone Survey Methodology*, New York, Wiley, pp. 509–28.

Ellis, A. (1947): 'Questionnaire vs. Interview Methods in the Study of Love Relationships', *American Sociological Review*, **3**, 541–53.

Galtung, J. (1967): *Theory and Methods of Social Research*, London, Allen.

Groves, R. M. (1989): *Survey Errors and Survey Costs*, New York, Wiley.

Groves, R. M. and Kahn, R. L. (1979): *Surveys by Telephone: A National Comparison with Personal Interviews*, New York, Academic.

Herman, J. B. (1977): 'Mixed-mode Data Collection: Telephone and Personal Interviewing', *Journal of Applied Psychology*, **62**, 399–404.

Hochstim, J. R. (1967): 'A Critical Comparison of Three Strategies of Collecting Data from Households', *Journal of the American Statistical Association*, **62**, 976–89.

Hyman, H. (1954): *Interviewing in Social Research*, Chicago, University of Chicago Press.

Jansen, W. *et al.* (1990): 'Operationalization and Measurement: The Primacy of Objective Data'. In: J. J. Hox and J. de Jong-Gierveld (eds.), *Operationalization and Research Strategy*, Amsterdam, Swets.

Kiesler, S. and Sproull, L. S. (1986): 'Response Effects in the Electronic Survey', *Public Opinion Quarterly*, **50**, 402–13.

Kooij, L. R. (1988): 'Eenoudergezinnen in de Huisartspraktijk [Single-Parent Families in the Practice of the Family Doctor]'. Doctoral Dissertation, University of Leiden.

Körmendi, E. and Noordhoek, J. (1989): *Data Quality and Telephone Interviews*, Danmarks Statistics, Copenhagen.

Levine, M. V. and Rubin, D. B. (1979): 'Measuring the Appropriateness of Multiple-Choice Test Scores', *Journal of Educational Statistics*, **4**, 269–90.

Mokken, R. J. (1971): *A Theory and Procedure of Scaling; With Applications in Political Research*, Den Haag, Mouton.

Nicholls II, W. L. and Groves, R. M. (1986): 'The Status of Computer Assisted Telephone Interviewing', *Journal of Official Statistics*, **2**, 93–115.

Perlman, D. and Peplau, L. A. (1981): 'Towards a Social Psychology of Loneliness'. In: R. Gilmour and S. W. Duck (eds.), *Personal Relationships in Disorder*, London, Academic, pp. 31–56.

Rogers, T. F. (1976): 'Interviews by Telephone and in Person: Quality of Response and Field Performance', *Public Opinion Quarterly*, **40**, 51–65.

Saris, W. E. *et al.* (1987): *A New Approach to Panel Research: Teleinterviewing*, Amsterdam, Sociometric Research Foundation.

Siemiatycki, J. (1979): 'A Comparison of Mail, Telephone and Home Interview Strategies for Household Health Surveys', *American Journal of Public Health*, **63**, 238–45.

Sijtsma, K. (1988): *Contributions to Mokken's Item Response Theory*, Amsterdam, Free University Press.

Sikkel, D. (1987): 'Computergestuurde Gegevens Verzameling', *Mens en Maatschappij*, 3, 289–301. [An English version of this article appeared in Sikkel, D. (1987): Quality Aspects of Statistical Data Collection', Amsterdam. Sociometric Research Foundation.]

Smitsmans, H. (1986): 'Eenzaamheid als Gezondheidsprobleem: Een Rangordening van Interventies en Voorzieningen [Loneliness as a Problem of Health: A Ranking of Interventions]'. Doctoral Dissertation, Health Sciences, Capaciteitsgroep GVO, University of Limburg.

Sudman, S. and Bradburn, N. M. (1974): *Response Effects in Surveys*, Chicago, Aldine.

Sudman, S. *et al.* (1965): 'The Effectiveness of Self-Administered Questionnaires', *Journal of Marketing Research*, **11**, 293–7.

Sykes, W. and Collins, M. (1989): 'Effects of Mode of Interview: Experiments in the U.K.' In: R. M. Groves *et al.* (eds), *Telephone Survey Methodology*, New York, Wiley.

Sykes, W. and Hoinville, G. (1985): *Telephone Interviewing on a Survey of Social Attitudes: A Comparison with Face-to-face Procedures*, London, Social and Community Planning Research.

Tatsuoka, K. K. and Tatsuoka, M. M. (1982): 'Detection of Aberrant Response Patterns and their Effect on Dimensionality', *Journal of Educational Statistics*, **7**, 215–31.

Tatsuoka, K. K. and Tatsuoka, M. M. (1983): 'Spotting Erroneous Rules of Operation by the Individual Consistency Index', *Journal of Educational Measurement*, **20**, 221–30.

van der Flier, H. (1980): *Vergelijkbaarheid van Individuele Testprestaties* [Comparability of Individual Test Scores], Lisse, Swets and Zeitlinger.

van der Flier, H. (1982): 'Deviant Response Patterns and Comparability of Test Scores', *Journal of Cross-Cultural Psychology*, **13**, 267–98.

van der Zouwen, J. and de Jong-Gierveld, J. (1987): 'De Vragenlijst in het Sociaal Onderzoek: Recente Toepassingen, Problemen en Veranderingen [The Questionnaire in Social Research: Recent Applications, Problems and Changes].' In: J. de Jong-

Gierveld and J. van der Zouwen (eds.), *De Vragenlijst in Sociaal Onderzoek: Een Confrontatie van Onderzoekspraktijk en -methodiek*, Deventer, Van Loghum Slaterus, pp. 11–22.

van Doorn, L. J. (1987): 'Het Gebruik van Microcomputers in Panelonderzoek [The Use of Microcomputers in Panel Research]'. In: *Recente Ontwikkelingen in het Marktonderzoek*, 1987–88 Yearbook of the Dutch Association of Market Researchers, pp. 9–21.

van Tilburg, T. G. (1988): 'Verkregen en Gewenste Ondersteuning in het Licht van Eenzaamheidservaringen [Obtained and Desired Social Support in Association with Loneliness]'. Doctoral Dissertation, Vrije Universiteit Amsterdam.

Wiseman, F. (1972): 'Methodological Bias in Public Opinion Surveys', *Public Opinion Quarterly*, **36**, 105–8.

Dr T. G. Tilburg is a KNAW fellow at the Department of Social Methodology, Vrije Universiteit Amsterdam.

Dr E. D. de Leeuw is a NWO fellow at the Department of Social Research Methodology, Vrije Universiteit Amsterdam, affiliated with the Interuniversity Institute for Psychometrics and Sociometrics (IOPS).